DOCUMENT RESUME

ED 423 786 HE 031 634

AUTHOR Jacobsen, Michele
TITLE Instructional Quality, Student Satisfaction, Student Success, and Student Evaluations of Faculty: What Are the Issues in Higher Education?
PUB DATE 1997-08-00
NOTE 33p.
PUB TYPE Information Analyses (070)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Faculty; *College Instruction; *Educational Quality; Educational Technology; Foreign Countries; Higher Education; Models; *Student Evaluation of Teacher Performance; Teacher Characteristics; *Teacher Effectiveness; Validity
IDENTIFIERS *University of Calgary (Canada)

ABSTRACT

This paper reviews issues in student satisfaction and evaluation of faculty as factors in determining instructional quality. It is based on a recent controversy at the University of Calgary (Alberta) which centered around the proposed publication of the results of student evaluations of faculty (SEFs) and was intended to help students, as consumers, make course choices. The paper then addresses seven topics related to the use of SEFs to assess instructional quality at the higher education level. These are: (1) the goals and values of higher education; (2) the educated person; (3) teaching effectiveness; (4) student satisfaction and student success; (5) the consumer model of higher education; (6) research on SEFs; and (7) technology and teaching. The paper notes that research indicates that student satisfaction can be influenced by such variables as teacher expressiveness, gender, rank of professor, expected grade in course, and possibly the integration of technology into instruction. The paper concludes that student satisfaction and student success cannot be separated from an assessment of instructional quality, but that a consumer model is detrimental to the goals of higher education. (Contains 82 references.) (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 786

Instructional Quality, Student Satisfaction, Student Success, and Student Evaluations of Faculty: What are the Issues in Higher Education?

Dr. Michele Jacobsen
University of Calgary
Calgary, Alberta, CANADA

Abstract

Those in higher education question whether the assessment of instructional quality can be separated from student satisfaction and student success, and how and why this might be done. In the process of exploring these questions, those in higher education must consider what likely went into the determination of an excellent education in the pre-computers in education era, and how those differences, if any, may or may not be an issue in today's campus classrooms. A recent controversy about student evaluations of faculty (SEFs) at the University of Calgary frames the following exploration of the multi-faceted and complex assessment and evaluation issues with regard to SEFs. Several topics are addressed in this paper: 1) the goals and values of higher education, 2) the educated person, 3) teaching effectiveness, 4) student satisfaction and student success, 5) the consumer model of higher education, 6) research on SEFs, and 7) technology and teaching. An argument is made that student satisfaction and student success cannot be separated from an assessment of instructional quality, but that a consumer model is detrimental to the goals of higher education.

Author Contact Information:

Dr. Michele Jacobsen
Department of Computer Science
University of Calgary
2500 University Drive, NW
Calgary, Alberta, CANADA
T2N 1N4

dmjacobs@ucalgary.ca
Office: (403) 220 4729
Fax: (403) 284 4707

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Michele Jacobsen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Introduction

A recent controversy shook the University of Calgary regarding student evaluations of faculty. Two main issues fueled the dispute: the implementation of a mandatory, universal teaching evaluation system, and the publication of results. The Students' Union (SU) advocates universal evaluations and publication of results to "improve teaching" (Jepson, 1996a). Although course evaluations have been mandatory at the U of C since 1992, the development and administration of instruments has been the responsibility of individual departments. Administrations' use the results to gauge faculty's teaching ability, but results are not available to students. The first issue, the adoption of a universal evaluation instrument, concerns faculty who believe that departmental distinctions warrant specialized student evaluations. Anton Colijn, president of The University of Calgary Faculty Association (TUCFA), worries that a set of universal questions will not adequately address faculty-specific issues (Skierka, 1997). According to a survey of TUCFA members, 55 percent of faculty oppose a universal instrument, while 72 percent oppose publication of results (Gazette, Jan 27, 1997). Despite faculty concerns, a 12-item instrument was recently approved on a pilot-test basis (Gazette, Feb. 24, 1997). The second issue, the publication of results, raises concerns about the interpretation of numerical ratings and the possible impact on instructional quality. Faculty question the validity of rating instruments and the reliability of student opinion. Colijn believes that "a number of factors could influence the ratings--such as an instructor's gender or race, how 'cool' the course subject is, the difficulty level of a course and whether the course is required or optional" (Gazette, Oct. 21, 1996) and cautions that "evaluation results have to be interpreted carefully" (Jepson, 1996b). Dr. Stalker, law professor and TUCFA Vice-President, said "many professors oppose the idea of releasing the results of student ratings--not as a matter of personal interest--but out of fear for the quality of education" (Gazette, Mar 24, 1997). The department head for Mathematics and Statistics agrees that published ratings may change how people teach, since instructors may change their style to improve ratings and popularity rather than actual teaching methods (Skierka, 1996a). Many fear that faculty will avoid teaching innovations that may negatively affect their course evaluations.

Students, on the other hand, are demanding information about teaching effectiveness for course selection. Eighty per cent of voting students in the Fall 1996 by-election were in favor of published results (Jepson, 1997). Paul Galbraith, (then) SU Vice-President, stated that "publishing evaluations allow students to tailor their choice of instructor to their learning style" (Skierka, 1996a). Much of the student rhetoric is based on a consumer model of education. The SU states that students are paying over $3,000 a year to attend the U of C and should be able to access information to help them choose an instructor (Gazette, Oct. 21, 1996). Students also believe they can judge good teaching, and that it is "important for students to know what other students think of the teachers...I think if we can make it into university we can understand how to rate teachers" (Skierka, 1996a). Although a number of issues remain to be resolved, the publication of instructor ratings has been endorsed in principle by the General Faculties Council (Gazette, Mar 24, 1997) and results may be available as early as Fall 1997.

The controversy over student evaluations of faculty (SEF) is not new, nor is it unique to the University of Calgary. In fact, this issue has been discussed and debated in the literature for more than half a century (for reviews, see Abrami, d'Apollonia, & Cohen, 1990; Costin, Greenough, & Menges, 1971; Cohen, 1981; Feldman, 1978, 1983; Marsh, 1982, 1984, 1994, 1995). On many campuses, SEFs are the sole measure of teaching effectiveness used for making tenure, promotion, retention, and merit decisions. Some faculty are uncomfortable that students have this much influence on their careers. Ryan, Anderson, and Birchler (1980) found that faculty believed that mandatory student evaluations affected their own and their colleagues' morale and job satisfaction when used for personnel decisions. At the same time, a consumer-oriented and politically active student population demands high quality educational products and services for their tuition dollars and access to collective consumer reports (i.e., SEFs) to guide their purchasing decisions. Included in this issue are the seemingly conflicting goals of a university. In order to attract sufficient enrollment, a university must maintain an excellent reputation for both research and teaching. However, faculty are often trained and selected for their research capabilities, not their teaching experience. As an initiation to their teaching role, new faculty are often assigned to

large enrollment introductory courses which are considered a challenge even by experienced professors (Sternberg, 1997). Students will not tolerate inadequate instruction as they pursue career goals which require not only a degree, but a very high GPA in order to compete. A final twist worth considering is the mutually dependent assessment roles of faculty and students. Faculty assess student success in their courses and report their findings in the form of grades. Students indicate their satisfaction with instruction by their responses to evaluation instruments at the end of term. Stone (1995) suggests the two assessment roles interact and influence each other: the higher the student's grade, the higher they rate the instructor. Ryan, Anderson, and Birchler (1980) report that a large percentage of faculty admit to lowering their academic standards to improve student ratings. Stone (1995) also suggests that rampant grade inflation is a result of enrollment driven funding that provides an incentive for increased bodies at the expense of academic standards. His thesis is based on the increased number of marginal or less prepared students being recruited by higher education institutions to fill available seats.

As demonstrated above, the use of SEFs to assess instructional quality in higher education is a complex, multi-faceted issue. The discussion that follows explores several topics that are relevant to understanding why SEFs are such a widely discussed topic in the literature: (1) the goals and values of higher education; (2) the educated person; (3) teaching effectiveness; (4) student satisfaction and student success; (5) the consumer model of higher education; (6) research on SEFs; and (5) technology and teaching. An argument will be made that student satisfaction and student success cannot be separated from an assessment of instructional quality, but that a consumer model is detrimental to the goals of higher education. Consideration of the primary goals and values of higher education and what it means to be an educated person will provide a conceptual framework for a discussion of research findings on the use of student evaluations of faculty. This paper will conclude with a discussion of information technology and teaching, and the potential implications of a technology-based curriculum for the assessment of an excellent education.

Goals and Values of Higher Education

An assessment of instructional quality in higher education must necessarily start with an *apriori* consideration of its basic goals and values. Most can agree that, despite the wide diversity of Canadian higher education institutions, the system has three main goals: education, research, and community or public service. Institutions tend to prioritize these goals differently according to their individual mandate or purpose. For example, universities tend to emphasize research, while vocational institutions tend to emphasize training and colleges tend to emphasize service. Further, the main goals may be differentially weighted across different faculties and departments within an institution, such as a university, and may often be seen as competing with each other. However, there are many ways to regard them as complementary and even mutually inclusive. Thus, excellent education and research are also important forms of community service, conducting research on teaching and learning is a way to enhance teaching, and effective teaching contributes to the development of skilled researchers. Recognizing that different priorities may drive an institution's primary mandate, all types of higher education institutions share a common commitment to their education function. Indeed, the very fact that these institutions fall under the jurisdiction of provincial and federal departments of advanced education signifies this shared goal of educating students. In order to fulfill their education function, universities must attract sufficient enrollment to help support the costs of fulfilling their other two functions: research and service. A university that maintains an excellent reputation for both research and teaching effectively improves their ability to attract students. Therefore, a university values its excellent reputation. Let us now consider how a university might evaluate its effectiveness in achieving educational goals.

The Educated Person

At first glance, it may seem relatively straightforward to determine whether a university is successfully accomplishing its educational goals. Thousands of university students are conferred degrees each year. However, careful consideration of the "educated person" reveals that the educational function of a university is more complex than just granting degrees. In his book examining the goals of an undergraduate education, Weingartner (1993) challenges us to consider

what it means to be an educated person. He presents one view that asserts in order for higher education to be effective, students who embark on the post-secondary path must start with the knowledge and abilities that enable them to march successfully toward the educated state that merits a culminating degree. Viewed from this *a priori* lens, undergraduate education is socially elitist, since specification of the outcome determines who is in a position to embark on this educational journey in the first place. Weingartner (1993) subscribes to an alternate, *a posteriori* view, which regards the student as possessing a complex capacity to learn; education is the process that actualizes this potentiality. Progress is made along a plurality of dimensions and is measured by what has been added to what existed before - by what has been learned. At its root, his a posteriori view is populist and the results sought by higher education constitute a broad range of achievements; it does not have a unitary design. In many ways, his educational philosophy is appropriate for Canadian higher education, and is hospitable to a heterogeneous population that is the product of huge waves of immigration from all over the world. Moreover, social progress depends on an educational philosophy that stresses above all the knowledge and skills that have been acquired, rather than simply the condition or certification that has been achieved. An a posteriori and populist view is not necessarily held by all who discuss the topic (Weingartner, 1993). A criticism of this view is the philosophy underlying professional education. This "production" approach to describing the goals of undergraduate education reasons that the educated person is a *product*; a competent physician, lawyer, teacher, or engineer. This line of reasoning ignores the role of formidable institutions that provide the setting for debates about the competent professional: guilds, licensing agencies, professional societies. Moreover, important criteria for conceptions of such professions are derived from their functions: healing the sick, defending the accused, teaching youth, building bridges, and so on. Weingartner (1993) asserts that being an educated person is not a profession; society does not delegate the job of defining that state to some expert organization. Nor is the function of an educated person limited to practicing a trade during working hours. Instead, he echoes Aristotle's formulation that being an educated person consists of living well for a whole life. Different philosophies and cultures will generate different accounts

of living well, and therefore different views of what an educated person is. Only in a society that is vastly more homogenous than ours could one expect to achieve a consensus or an overarching, unitary view on so fundamental a set of issues. A broad characterization of the goals of undergraduate education can heighten our awareness that universities are not just course-giving institutions; effectiveness in education, rather, requires one to be open to a complex set of pedagogic means that are or can be available in institutions of undergraduate education.

An additional consideration is that we are, above all, examining the education of adolescents. Although the age of first-year students is increasing, the largest proportion of undergraduates is composed of pre-twenty year olds who begin university shortly after high school. Weingartner (1993) asserts that we cannot assume that students entering university are automatically motivated to act in ways that will make them good students. Indeed, if aptitude and strength of motivation were made the sole criteria for selection into university, beyond past high school performance, the ranks of undergraduates might be thinned considerably. Because aptitude and motivation are not seriously considered in admission decisions, one is left with a need to be concerned with the motivation of students in the pedagogy of undergraduate education. A professor who believes that motivation is solely the students' concern would probably be very pleased with the smaller, "better" classes that would result if motivation were used as an admissions criterion. The drawback, however, is that far fewer professors would be needed to teach a reduced number of undergraduates, possibly leaving that very professor without a job.

The fact that motivation is relevant to undergraduate education, and the suggestion that the majority of undergraduates are yet to become fully formed adults, implies that undergraduate institutions are engaged in the business of shaping character (Weingartner, 1993). In the sense that growth and maturation continue during the years of undergraduate study, what happens during those years influences the direction of those developments. Viewed from this perspective, universities have an obligation to try to give shape to this direction, in consonance with their educational goals. That higher education should aim at fostering in its students those qualities of mind that were classically known as "intellectual virtues" has not been controversial. Sometimes

referred to as building a knowledge base, individual programs of study include instructional objectives ranging in levels of understanding from proficiency, conversancy, competency, to mastery, with outcomes ranging from measurable to difficult to measure. A comprehension level goal may be to demonstrate proficiency in the use of the French language, for which the design of reliable tests is not difficult. Conversancy, which may require analysis and synthesis, would involve having a grasp of French literature, which is more difficult to measure. We can and do design trustworthy tests to determine whether a student is proficient at expository writing or applying formulae to solve problems. Aiming at higher goals of conversancy and competency are more distinctive goals of undergraduate education than is the teaching of proficiencies. The qualities of mind referred to here include, but are not limited to, the ability to suspend judgment in the face of inadequate evidence, motivation to invest oneself in learning, or to the capacity to remain open to the views of others. These "intellectual virtues" are not only skills, like the ability to solve algebraic equations, nor are they knowledge, like an understanding of Einstein's Theory of Relativity. Instead, they are an indication of the depth of one's character. However, there is no feasible way of examining for the capacity to make courageous moral decisions (Weingartner, 1993). Solid evidence for this kind of intellectual growth is not routinely available at the conclusion of every semester. Hence, one could find faculty who would disagree that character development is their responsibility because it is so difficult to measure. However, as we shall see, expert professors readily accept this responsibility and challenge.

Teaching effectiveness: A complex set of pedagogic goals and means

Having considered the overall goals of an undergraduate education, let us now turn our attention to the views of individual faculty who have been identified as expert teachers and what they believe constitutes an "excellent university education". Faculty often acknowledge that graduate school may not have ideally prepared them for their teaching role. However, judging from the sheer volume of books (Andre & Frost, 1997; Bess, 1997; Cahn, 1978; Ellis, 1993; Flood & Moll, 1990; Johnson, 1995; Jones, 1995; Kimball, 1988; McKeachie, 1978; Sternberg, 1997) and journal articles published on teaching excellence in higher education, the majority of faculty are

highly motivated and committed to improving instructional quality. Sternberg's (1997) book profiles the views of expert professors who have not only taught introductory psychology, but have written textbooks for the course. There is remarkable consistency in what the individual professors regard as important and valued instructional goals. They want to portray their discipline as: (a) an empirical science based on critical thinking, (b) knowledge that is dynamic, not static, (c) a broad and diverse discipline, and (d) important in everyday life. These goals could easily be applied to other disciplines. David Myers explains his teaching philosophy in this way: "What greater life mission would one hope for than to do one's part to restrain intuition with critical thinking, judgmentalism with compassion, and illusion with understanding?" (p 109). Each professor describes a complex set of pedagogical means they use in their attempt to ignite a fire in students that is fueled by an enthusiasm and excitement about psychology. They want to convey their passion for the subject and consider it a measure of their success when students go on to pursue graduate studies or even professional careers in psychology. An important means to consider is establishing the "benchmark" or challenge that many professors view as crucial to developing a student's learning capabilities and their character. Excellent professors aim to challenge their students to grow intellectually and strive for greater depth of understanding. Sternberg (1997) takes the position that professors have a responsibility to teach to a student's weaknesses as well as to their strengths. He believes that students need to be taught to think in at least three different ways: analytically, creatively, and practically (p. 141), and acknowledges that general intellectual ability means that almost everyone is good at some things and not so good at others. A student may be strong analytically, but not particularly strong creatively, and so on. Therefore, a teacher who aims to improve analytical, creative, and practical thinking will practically ensure that they are reaching some of the students some of the time, while not reaching all of the students all of the time. Many courses emphasize one particular type of thinking, and students who already think that way do quite well. Sternberg (1997) believes that in order for students to be successful in later life, they need to learn their own individual pattern of strengths and weaknesses, and adopt strategies to compensate for and correct their weaknesses. Therefore, a worthy goal of

undergraduate education is to intellectually challenge students to modify and increase their abilities, not just to practice what they already know how to do well. In this way, Sternberg (1997) believes instruction prepares students for the many kinds of intellectual challenges they will encounter throughout their lives. Recall the Student Union's argument that published student ratings will aid students in selecting an instructor that matches their learning style. If we agree that Sternberg's (1997) goal to challenge students to address their learning weaknesses is a valued endeavor in undergraduate education, and that professors are in the business of shaping character to prepare students to live well for a whole life (Weingartner, 1993), we have to wonder whether this "consumer demand" is actually in the best interests of students. One could argue that students have the right to limit their learning experiences to only those in which they believe they will excel. After all, students are driven to achieve the highest GPA possible to better their chances for employment, program selection, or graduate school admission. However, if we also agree that universities are not just degree granting institutions, then we must believe that challenging students to grow beyond their present strengths and develop new ways of thinking is a goal worth preserving. The use of published ratings so that students can match their learning style to that of the professor may inadvertently reduce the value and outcome of the undergraduate education that students' receive.

Student Satisfaction and Student Success

How important is it to include measures of student satisfaction and student success in the assessment of the instructional quality of an undergraduate education? Astin (1993) asserts that assessment results are of most value when they shed light on the causal connection between educational practice (means) and educational outcomes (goals). If the instructional goals of undergraduate education include the developing of a character as well as intellectual virtues, and preparing students to live well for a whole life, and the pedagogic means to achieve this are complex and demanding, then how do we assess whether we are achieving excellence? One present method for evaluating student learning is the course grade assigned at the end of semester. Another is to infer student satisfaction with instruction from student evaluations of faculty. To make an evaluation, a measurement is interpreted according to some value system. The quality of

the evaluation depends on both the quality of measurement and the care with which this result is interpreted. A careless interpretation of good quality data is likely to lead to a poor evaluation just as a careful interpretation of shoddy data would (Violato, McDougall, and Marini, 1992). Although one can argue that current methods for assessing student success and satisfaction are subject to various sources of bias and error, they are still our "best guess" about whether universities are achieving their instructional goals. An assessment of instructional quality cannot be separated from student's perceptions of the quality of instruction and student success. Student satisfaction is an important measure of teaching effectiveness: "If one takes increased interest and motivation for learning as important outcomes, it is hard to come up with better measures than the students' own perceptions of their interest. Not only can students provide data about the effects that instruction has had on them, but they also have an excellent opportunity to observe what the teacher does and what the course requires" (McKeachie, 1990, p. 194). It has been established that student motivation is an instructional goal of expert professors (Weingartner, 1993; Sternberg, 1997) and as such should be a considered a good indicator of teaching effectiveness. In addition, undergraduate education goals are achieved, and the curriculum actually delivered, only to the extent to which students learn (Weingartner, 1993). Professor assigned grades or other achievement measures are our "best guess" that learning has occurred. In subsequent sections, the extent to which learning has occurred, and or conditions under which learning can and does occur, will be examined with respect to student satisfaction with the teaching and learning process.

Consumer Model of Higher Education

However important student satisfaction is when assessing instructional quality, this author agrees with critics who are against the often used "consumer student" rationale for publishing the results of SEFs. Haskell (1997) suggests we must consider the "consumer" of higher education to be society as a whole, rather than limiting consideration of the impact of an undergraduate education to the individual student. The financial argument supporting the consumer student's demand for their rights is on shaky ground when one considers that Canadian university students pay only a small proportion of the true cost of their education. In large part, the student's education

is funded by tax dollars allocated by provincial and federal departments of education. Society collectively invests in higher education because of the predicted benefits to society as a whole.

One of the arguments against the consumer model is founded in the belief that SEFs endanger academic freedom. Schrank (1993), Stone (1995) and Haskell (1997) believe that because SEFs are used for tenure, promotion, retention, and merit decisions, faculty may feel pressured by administration and students to modify what they teach about in order to "avoid trouble". Schrank (1993) argues that academic freedom is important within the liberal conception of a university because universities are not only purveyors of the known, but creators and disseminators of the new. To fulfill this crucial creative role, universities require openness, the free exchange of ideas, controversy, ferment, dissent, even heresy. Schrank (1993) believes a component of academic freedom is tolerance, and that participating in the university environment may sometimes involve accepting a fair measure of intellectual discomfort. The current debate on political correctness and the suppression of speech on the ground that it expresses harmful or hurtful ideas is incompatible with the fundamental premises of an intellectual community. This argument is not an attempt to protect the "rights" of a faculty member who was racist or sexist. However, one must recognize the ambiguity of these terms, and the dependence on interpretation by both the purveyor and the offended party. Schrank (1993) suggests the appropriate response to speech expressing ideas that are regarded as false, politically unacceptable or dangerous is speech that refutes those ideas. Haskell (1997) and Stone (1995) argue that threats to academic freedom come from within an institution in the form of the "consumer student". They argue that within a consumer model of education, to deny consumer demands is difficult and often results in a watered down curriculum and inflated grades. Further, SEFs are seen as fiscally efficient way to determine who is serving the "consumer student's" demands best. To deny consumer demands in this type of environment is viewed as undemocratic and downright mean spirited. However, Haskell (1997), Stone (1995) and McMurtry (1991) argue that an attempt to satisfy the consumer student's demand for their political correctness or a high GPA educational product is problematic, if not detrimental, to the goals of higher education.

McMurtry (1991) forms a clear and convincing argument against the utility and validity of a consumer model of higher education, and the explicit shaping and forming of higher education goals and purposes to supply the demands of a student-consumer market.

> The best product on the market, as we know, is the one which is the most 'problem-free' for its purchaser--delivered ready made for 'instant easy use', 'guaranteed replacement' it if does not work, and 'repaired cost-free' whenever it needs maintenance attention. The best education, on the other hand, is the opposite on all standards of excellence. It cannot be produced or delivered by another at all, is never ready-made nor instant, and cannot be guaranteed replacement or service cost-free if it is not working. The higher the standards it has, the less it can be immediate in yield, the more work it demands of its owner, and the more its failures must be overcome by its possessor's own work. An education can never be 'problem free', and poses ever deeper and wider problems the higher the level of excellence it achieves (p 213).

It follows, therefore, that if the goals of undergraduate education, such as the development of character, motivation, and intellectual virtues, and the creation and dissemination of new ideas, are subverted to the production model of certification in a consumer driven market, then universities will lose their ability to contribute in meaningful ways to society. Students cannot expect to passively receive an excellent education, nor can professors be expected to give one. One of the disadvantages of referring to students as customers is that it obscures their central role as producers of knowledge, in this case their skills and knowledge of the world and their discipline (Mingle, 1995). The development of an educated person who will live well for a whole life must be a shared endeavor between faculty and students. Both parties must have a say in the quality of this transaction, and no one party should have sole responsibility for defining and directing the efforts of the other. Students cannot expect to drive the curriculum to suit their desire for certification. Faculty, as recognized experts in their discipline, must take seriously their responsibility to teach well and share their enthusiasm and knowledge with students. As apprentices in a discipline, students must afford faculty the rights and responsibilities that come with expertise, and become willing partners in the exploration of a discipline. In order to protect societies' investment in higher education, faculty need to continue to set high standards for conversancy in a discipline while motivating students to engage in the learning process. Students need to provide thoughtful feedback on their satisfaction with the instructional process and the quality of teaching.

Research on Student Evaluations of Faculty

Two basic premises underlie assessment in higher education (Asten, 1993). One, an institution's assessment practices are a reflection of its values. A university values its reputation for excellent teaching, and therefore collects and pays attention to information about instructional quality. The second premise is that assessment practices should further the basic aims and purposes of our higher education institutions. The basic motive for gathering information about the quality of instruction should be to improve the functioning of its institution and its people. It is important to keep these premises in mind when considering the use of information gathered by SEFs. First, because universities rely on SEFs to evaluate faculty teaching for merit, tenure, promotion and salary decisions, it is important to assure that they provide valid indications of instructional quality. Second, because results from student ratings are expected to help improve instructional quality, then it is important to examine what mechanisms are in place to capitalize on the information obtained. The dispute over publication of results at the University of Calgary revealed a number of faculty beliefs regarding student evaluations. Arreola (1995) provides a good framework for exploring faculty beliefs about SEFs. Similar categories of beliefs about SEFs will be used to organize the following selective review of current and past research findings on SEFs. This section will include a consideration of the use of SEFs to improve instructional quality.

Students as unqualified and impulsive judges

One common belief is that students cannot make consistent judgments about the instructor and instruction because of their immaturity, lack of experience, and capriciousness. Costin, Greenough, and Menges (1971) and Hogan (1973) found correlations between student ratings of the same instructors and courses over time ranged from 0.70 to 0.87., so this belief can be disregarded as essentially untrue. It appears that students are less impulsive than generally believed, and that ratings across instructors and courses, and even over time, are fairly reliable. The first belief is based on an assumption that only colleagues are qualified to evaluate their peers' instruction, Researchers have explored the relationship between student and peer ratings of teaching with variable results. Maslow & Zimmerman (1956) found that student's judgments of

their teacher's effectiveness correlated highly (r = 0.69) with faculty judgments of the same teachers. Doyle and Crichton (1978) compared student, peer, and self evaluations of college instructors on overall teaching ability. Significant correlations were found between student and peer evaluations (r = .49), student and self evaluations (r = .52), and colleague and self evaluations (r = .64), which provides support for the consistency of student evaluations when compared with other "expert" opinions of teaching ability. Kremer (1990) found that the overall reliability of peer ratings varied depending on the professional area being assessed. Interjudge reliability of peer ratings of research (r = .77) were highest, followed by service (.56) and teaching (.50). Peers are more confident rating a peer's research ability, rather than service or teaching performance. Rather than supporting the belief that peer's are better judges of teaching than students, the research cited above supports the use of *both* student and peer evaluations of teaching ability.

Murray, Rushton, and Paunonen (1990) provide an interesting perspective on what peer and student ratings can tell us about teaching effectiveness in different types of courses. They investigated relations between peer ratings of faculty personality traits and student ratings of teaching effectiveness in six types of university psychology courses (from introductory non-major to graduate). They found clear evidence that perceived teaching effectiveness varies substantially across different types of courses for the same instructor. Correlations between cumulative instructor ratings in different types of courses ranged from .06 to .78 and averaged .49, indicating that teaching effectiveness shows only moderate cross-situational consistency. It is the exception rather than the rule for an instructor to perform exceptionally well or exceptionally poorly in all types of courses. The authors posit that although faculty are trained and selected for their more homogenous cognitive and intellectual characteristics, it is their heterogeneous personality characteristics that best predict teaching effectiveness. Because personality characteristics, as measured by peers, may predict teaching effectiveness in various types of courses, as measured by student evaluations, then departments should assign teaching tasks according to the type of situation in which faculty are most likely to excel. Benefits from this arrangement would accrue to departments, faculty and students, as well as increase an institution's reputation for good teaching.

Some faculty believe that students are unable to make accurate judgments until they have been away from the course, and possibly away from the university, for several years. Drucker and Remmers (1951) showed that alumni who have been out of school from five to ten years rated instructors much the same as students currently enrolled. Recent evidence by Overall & Marsh (1980) and Marsh (1984) further substantiates this earlier finding, and shows this belief to be essentially untrue. Students' evaluations collected at the end of a course are remarkably similar to the retrospective ratings provided by the same students several years later.

Student Ratings as a Popularity Contest

Faculty do not want to believe that enthusiasm or expressiveness is all that is needed for one to be perceived as an effective teacher. As a result, teacher expressiveness, which includes enthusiasm, humor, friendliness, physical movement, vocal inflection, warmth, and charisma, has been examined for its effect on student ratings of faculty. Perhaps the most famous set of experiments in this area are the Dr. Fox studies on the "educational seduction" phenomena. Ware and Williams (Ware & Williams, 1975, 1977; Williams & Ware, 1976, 1977) and Meier & Feldhusen (1979) conducted laboratory studies of student's reactions to video-taped lectures presented by a professional actor. The lectures differed on two variables: instructor expressiveness and amount of lecture content. After viewing the tape(s), students rated the instructor and the content, as well as completing a quiz on the lecture material. General results indicated that: (a) more substantive lecture content results in greater student achievement than less substantive lecture content, (b) student ratings reflect the amount of lecture content when instructor enthusiasm is low, (c) student ratings do not reflect the amount of lecture content presented when instructor enthusiasm is high, and (d) a highly expressive presentation will receive higher student ratings that will a presentation that lacks expressiveness, regardless of the amount of substantive content presented. These findings were substantiated by similar laboratory research (Bascow & Distenfeld, 1985; Basow, 1990) which also found significant main effects for teacher expressiveness over content. Perry, Abrami, & Leventhal (1979) conducted a "Dr. Fox" type study using videotapes of a real psychology professor, but failed to replicate all of Williams & Ware's (1976, 1977)

results. However, they did replicate the earlier finding that regardless of content (high vs. medium), expressiveness affected student ratings of lectures more than achievement, and conversely, in low expressiveness lectures, content affected ratings more than achievement. This suggests that students taught by instructors lacking enthusiasm, humor, friendliness, charisma, and the like may rate their instructors more severely on the basis of lecture content than their actual achievement would justify. When students are unsatisfied with instructional delivery, they rate the instructor lower regardless of how well they succeed in the course.

The Dr. Fox studies have been criticized for lack of generalizability because the experiments were not situated in natural classroom settings. However, a recent study by Stephen J. Ceci ("Student ratings news", 1997; Murray, 1997) suggests that the "halo-effect" for expressive instructors, regardless of content, carries over to the classroom. In this study, Ceci taught his developmental psychology course two semesters in a row using the same syllabus, films, textbook and tests. The second time he taught the course, Ceci varied his presentation style by varying his pitch, speaking more enthusiastically, and using more gestures. Unsurprisingly, student ratings of his enthusiasm improved. However, a surprising difference in student evaluations was that students rated Ceci's level of knowledge, organization, fairness, and even the quality of the textbook much higher in the second course. Student achievement on tests was basically the same in both semesters. Ceci suggests that "student ratings are far from the bias-free indicators of instructor effectiveness and quality that many have touted them to be" (p. 5). Ceci's findings support the proposition that student perception can be manipulated by teacher expressiveness.

In another classroom-based study, Widmeyer & Loy (1988) found that a visiting lecturer who was introduced as either "warm" or "cold" was rated significantly differently by students immediately after the presentation. Students who were told the lecturer was a "warm" person rated him as more pleasant and sociable, less irritable and ruthless, more humorous, less formal, and more humane than did students who were told that the same guest lecturer was a "cold" person. Student ratings of the visiting lecturer's teaching ability were also affected, with the "warm" lecturer receiving significantly higher ratings. This study supports the contention that positive

attributions to a stimulus person greatly influences subjects' overall impressions of personality and teaching ability. Findings on teacher expressiveness suggest that if instructors want to positively affect how students rate their personalities and teaching ability, they should present themselves as "warm", speak enthusiastically, and use more gestures. A sense of humor also improves a teacher's ratings. Gorham & Christophel (1990) investigated the relationship between a teacher's use of humor and overall immediacy and student's perceived cognitive and affective learning outcomes. The total number of humorous incidents recorded for each teacher was positively correlated with the frequency of his/her use immediacy behaviors, and the overall use of immediacy behaviors was highly correlated with students' perceived learning outcomes. Bryant, Comisky, Crane, & Zillman (1980) also found a strong correlational relationship between teachers' use of humor and students' evaluations of their teachers' effectiveness.

Instructor warmth, humor, and immediacy, all of which are elements of "expressiveness", seems to positively affect student's ratings of instructors and their perception of engagement in the course. Perhaps the more expressive instructor is not just "more popular" but is actually constructing a classroom environment more conducive to engagement, which may lead to increased learning. Additional evidence that supports this proposition comes from a study conducted by Frymier and Thompson (1992) that explored associations between students' motivation to study, teachers' perceived credibility, and teachers' use of affinity-seeking strategies. Results indicated that several affinity-seeking strategies (i.e., interpersonal skills such as listening, optimism, sensitivity, openness, facilitating enjoyment, trustworthiness, comfortable with self, dynamism and supportiveness) were significantly associated with competence and character, indicating that the use of affinity-seeking in the classroom may assist in the development of teacher credibility. Students' perceptions of teacher credibility and teachers' use of affinity-seeking strategies were found to be positively and significantly associated with students' motivation to study. The correlations here suggest that teachers who use student-oriented affinity-seeking strategies that indicate teacher interest in and respect for students are more likely to have students with higher levels of motivation, as well as be perceived as more credible (Frymier & Thompson, 1992). What

can be concluded from research on expressiveness is that the instructor who actively attempts to be more expressive, warm, humorous, affinity-seeking and immediate will have positive effects not only on student ratings of their teaching, but may construct classroom climates more conducive to motivation and engagement. Although Ceci did not find a difference in achievement, the Dr. Fox studies found that more substantive lecture content resulted in better student achievement. It seems that the more expressive teacher with substantive lectures will positively affect both student satisfaction and student success.

Gender effects

A common belief suggests that gender of the student and the instructor affects student ratings. Because university teaching is still a predominantly male profession, and SEFs have the potential to influence advancement in such a career, it is worthwhile taking a closer look at possible gender biases in SEFs. Gender has typically been examined in terms of whether male and female professors receive different mean ratings. Basow and her colleagues have looked at the issue of gender bias with a focus on more subtle effects (Basow and Distenfeld, 1985; Basow and Silberg, 1987; Basow, 1990; Basow, 1995). A complex picture emerges when SEFs are examined as a function not only of professor gender but also student gender, gender-typing of the discipline, status of the professor, and gender-typed characteristics of the professor (Basow, 1995). Male students have been found to rate female faculty lower than they rate male faculty, especially when faculty gender differences in rank, discipline, and student-perceived personality traits are controlled for (Basow & Silberg, 1987). However, when such variables are not controlled for, female faculty are sometimes rated higher than male faculty (Feldman, 1983). Different aspects of teaching also appear important for male and female professors in obtaining good overall ratings, and both male and female students have gender-related expectations for their professors (Bennett, 1982; Kierstead, D'Agostino, & Dill, 1988). For example, Sandler (1991) found that it seems more important for female professors to demonstrate warmth and friendliness than it is for male professors. With respect to sensitivity, student comfort, and respect, female teachers are rated higher than male professors, especially by female students (Basow, 1995). Bennet (1982) found

that a highly structured instructional approach was viewed by students as more professional, and when women faculty used a more collaborative and innovative method of teaching they were seen as less competent. Male and female faculty also may differ in their teaching style or may appear to students to possess different personality traits (Basow & Silberg, 1987). For example, female professors appear to run a more democratic classroom and generally are rated higher than their male counterparts in interpersonal traits and behaviors, such as helping students (Feldman, 1983; Murray, Rushton, & Paunonen, 1990). Basow & Silberg (1987) found that although women tended to rate their female professors somewhat higher than their male professors, they tended to view their male professors as more dynamic and as better teachers.

An examination of gender effects is further complicated by the fact that male and female professors are not similarly situated or represented on Canadian campuses (Table 1). Women are overrepresented in the lecturer positions and underrepresented in the ranks of the professoriat.

Table 1
Full-time Faculty by Gender, Type of Appointment, and Rank 1995

	Full Professors	Associate Professors	Assistant Professors	Lecturers
Female Contract	.09 %	.34 %	5.8 %	36.2 %
Male Contract	0.64 %	1.0 %	9.2 %	30.6 %
Female Tenured/Leading to Tenure	11.0 %	25.9 %	36.3 %	15.1 %
Male Tenured/Leading to Tenure	88.0 %	61.1 %	47.7 %	10.4 %

Source: Statistics Canada. Postsecondary Education Section. Unpublished data.

Students are much more likely to be taught by tenured, high ranking male professors during their university career, and less likely to be taught by women professors. Because higher evaluations tend to be associated with rank, women may lose out. For example, tenured professors tend to receive better ratings than untenured professors and teaching assistants; upper level courses tend to be rated higher than lower level courses (Feldman, 1983). Perhaps because of their lower representation on campus, female professors are seen as atypical and therefore their gender is notable. Basow (1995) found that male faculty were rated similarly by their female and male students, regardless of divisional affiliation, whereas female faculty were frequently rated differently by their female and male students, especially in the humanities and social sciences. Male

students frequently rated female professors the lowest on fairness, thought stimulation, and overall teaching. The practical significance of this research, despite somewhat small reported effect sizes, is that in the case of the untenured female faculty member with few years of teaching experience in the humanities or the social sciences, who teaches a large number of male students who expect low grades, these small effect sizes may add up to gender-biased perceptions and evaluations, as found in an earlier study (Basow & Silberg, 1987). Research demonstrates the complexity of gender variables when they operate in the university setting, but is by no means conclusive. Further investigation is warranted to better understand and predict the impact of stereotypical gender expectations on student evaluations in the predominantly male university teaching profession.

Grading effects

One common belief is that the grades or marks students receive in the course are highly correlated with their ratings of the course and the instructor. Stone (1995) believes instructors are rated higher when they lower standards and artificially raise grades, however, research on this question has yet to produce consistent and conclusive results. Abrami, Dickens, Perry, & Leventhal (1980) failed to find any evidence of a large grading-standards effect on ratings, and considered the size of the differences they did find to be relatively unimportant when student ratings are used to make gross distinctions between instructors. Others have found that increased grades or more lenient grading criteria do result in higher student ratings of teaching effectiveness. Holmes (1972) found that when students' grades disconfirm their expectancies (i.e., they believe they will receive an A but are told they received a B) they will tend to rate the instructor lower on teaching effectiveness. Powell (1977) found that evaluations of both the instructor's performance and the course decreased as the stringency of grading criteria increased. DuCette and Kenney (1982) found that correlations between grades and evaluations vary across types of courses, with required courses showing the strongest correlation between grades and ratings. Some researchers interpret these findings to suggest that higher grades are a natural result of more effective teaching. Howard & Maxwell (1980, 1982) and Beatty & Zahn (1990) suggest that the relationship between grades and student satisfaction is a welcome result of important causal relationships among other

variables, such as student motivation and engagement, rather than simply as evidence of contamination due to grading leniency. This line of research may also lead to investigations of the relationship between student assessment and instructional goals and means. Expert professors agree that assessment and instruction must agree (Sternberg, 1997). If students are to be tested for thinking but are instructed in ways that emphasizes the acquisition of facts, then students will be confronted with assessment that does not match the way in which they were taught. Students will quickly learn that they will not be held accountable for thinking. The cost of this mismatch will be a loss of students' trust in the instructor's credibility. Holmes' (1972) study on grade expectancy suggests that students' ratings are also an indication of their trust in the assessment practices of an instructor. A question worth asking with regards to grading effects, that is not dealt with in the literature, is whether a professor's grading practices, as perceived by students as a fair or unfair assessment of their performance, may affect ratings of the course. Two items on the universal ratings scale approved by the General Faculties Council at the University of Calgary (Gazette, Feb. 24, 1997) will measure student opinion about grading practices. Perhaps the pilot-study of this instrument should evaluate whether students' opinions about grading, as measured by responses to these two items, are related to measures of overall teaching quality. In conclusion, the belief that student ratings are highly correlated with their grades can be supported or unsupported by the literature, and is still an open question that warrants further investigation.

Validity, Reliability, Usefulness

Researchers have questioned the validity, reliability, and usefulness of student evaluation instruments on various dimensions using different research methodologies (for reviews, see Abrami, d'Apollonia, & Cohen, 1990; Costin, Greenough, & Menges, 1971; Cohen, 1981; Feldman, 1978; Feldman, 1983; Howard, Conway, & Maxwell, 1985; Marsh, 1982, 1984). Further meta-analysis of the research on the design and testing of valid and reliable instruments is not a goal of this paper. However, an examination of the usefulness of student evaluations with regards to whether they can be used to improve instructional quality is important to the present discussion. McKeachie et al. (1980) found little effect of ratings alone on teacher improvement, but

they found significant improvement when the ratings were communicated in a face-to-face counseling session. Marsh and Roche (1993) investigated the efficacy of a student evaluation process that included instructor self-evaluations and a feedback/consultation intervention. As a result of midterm student and self evaluations, faculty targeted specific aspects of their teaching with the help of the consultation advice from excellent instructors. The student ratings at end of term improved significantly for instructors who employed these techniques compared to a control group who did not. Marsh and Roche's (1993) findings suggest that student evaluation feedback coupled with expert consultation is an effective means to improve teaching effectiveness. In a similar study that employed feedback to faculty, Abbott, Wulff, Nyquist, Ropp, and Hess (1990) examined student satisfaction with eight processes of collecting student ratings of instruction by varying (a) method (group interviews vs. individual standardized ratings forms), (b) timing (midterm vs. end of course), and (c) amount of instructor reaction to student ratings (restricted vs. extended). Consistent with predictions drawn from reactance and social comparison theories, Abbott, et al. (1990) found that students were more satisfied with interview methods at midterm followed by extended instructor reaction than with traditional approaches for collecting student opinions about instruction (i.e., standardized ratings forms administered at the end of a course). The authors believe that students were more satisfied with the group interview process at midterm because they were able to compare their opinions about the course with others, and they expected to benefit more fully from changes the instructor instigated as a result of the feedback. Findings from this study indicate that students are most satisfied with a faculty and course evaluation process at midterm that enables them to observe the effects and benefit from their feedback on instruction. The various approaches to investigating the usefulness of student evaluations for improving instructional quality indicate that standardized instruments administered at the end of term have less impact on teaching effectiveness than midterm formative evaluations that provide feedback to instructors. If the Student's Union at the University of Calgary truly wants to help improve the quality of teaching on campus, then perhaps they should direct their lobbying efforts towards the implementation of a midterm course evaluation process that provides useful and timely

feedback to the instructor, rather than arguing for a process (standardized instrument administered at the end of course) that has been shown to have limited impact on teaching effectiveness.

Teaching and Technology

A relevant area to consider as part of the complex set of pedagogic means used to deliver undergraduate education is the use of instructional technology for teaching and learning. Universities are motivated to spend vast amounts of money acquiring technology because it increases their reputation for innovative teaching, and in order to prepare students to live well for a whole life in a technology-rich society, higher education must promote the acquisition of technology-related knowledge and skills across disciplines. Formal evidence linking this investment to student achievement and motivation has been modest, but encouraging (Kulik & Kulik, 1980, 1987; Ehrmann, 1995). Though enthusiastic claims for the impact of computer integration upon achievement have not been realized, the motivational aspects of technology have been recognized as a sufficient condition upon which to base their acquisition and use (McKeachie, 1990). A technology-based curriculum seems to be a win-win situation for both faculty and students. However, a belief expressed at the University of Calgary is that SEFs may cause some faculty to avoid teaching innovations or experiments that may negatively affect their evaluations. There is no research that specifically investigates the relationship between SEFs and the integration of technology. On the surface, faculty concern is hard to justify in light of findings that technology-based instruction made small but significant contributions to student success and student satisfaction. However, a possible limitation to relying on student recognition of teaching excellence and technology integration lies in the initial demands made on novice students and faculty, and the nature of changing faculty and student roles as a result of integrating technology.

Although technology has been increasingly available and used for teaching and learning since the microcomputer revolution of the 1970s, students still vary greatly in their readiness for participation in a computer-using portion of a syllabus, while faculty members are most unequally prepared to make use of computers in their teaching (Weingartner, 1993). The instructor who plans to require the use of computers must plan strategies to address both weak or non-existent computer

skills in themselves and students, and also develop tasks to appeal to those who have already mastered the needed skills. This "anticipatory remediation" (Weingartner, 1993) for both novice faculty and novice students may impact SEFs during the transition, as both teacher and learner grapple with a relatively new and complex means of instruction.

Integrating technology into the teaching-learning transaction has been found to transform the teacher's role from being the traditional "sage on the stage" to being a "guide on the side" as student roles change from being passive receivers of content to being more active participants and partners in the learning process (Roblyer, Edwards, & Havriluk, 1997). The transition from teacher-directed to more student-centered instruction has been the result of a move from predominantly behaviorist to more constructivist approaches to student learning. It was originally believed that technology would serve a behaviorist, "teaching-machine" role in the classroom (Skinner, 1954). Along with a transition in the work world from a mass production, industrial model of the physical laborer, to a model of the knowledge worker in an information age, there has been a transition in education and psychology from a behaviorist, stimulus-response-feedback model of the learner to a more constructivist, knowledge building, information processing model of the learner. Constructivists, like Seymour Papert, envisioned a new role for the computer, that of partner in the active knowledge building and problem solving learning processes of the student. In today's technology-supported undergraduate classrooms, it is hypothesized that students may initially resist becoming more actively involved in their own learning, because of past success with behaviorist and traditional approaches, and may actually react less positively to the change from teacher-centered, content-driven learning, to a more student-centered constructivist approaches that require more effort and engagement on part of the student. Faculty members may actually be integrating technology in effective ways to increase student involvement and engagement in their own knowledge construction, and the result may be that students react less positively to the increased workload and demand, and actually rate the faculty member less favorably than another faculty member who expects more traditional, passive forms of participation from students. Add to this the frustrations of hardware, software, and network instability, variable access, and the

newness of the field and experimental teaching methods, and the integration of technology for teaching and learning becomes more complex and complicated. This proposition has not been explored using experimental conditions, however, it certainly warrants further investigation to determine whether there are "growing pains" in the transition period while students and faculty struggle with technology, and whether this affects SEFs.

Conclusion

This paper proposed no major solutions to problems associated with SEFs. It was argued that an assessment of instructional quality cannot be separated from student satisfaction and student success. One of the primary goals of a university is to educate students, and the extent to which students learn is an assessment of its performance. It is the professor's responsibility to motivate and engage students in the learning process, while setting rigorous standards for achievement. Student satisfaction with instruction is an important variable to formatively assess in order to constantly improve instruction. Ratings of instruction can be effective in improving instructional quality when collected mid-way during the semester and discussed with the faculty member. It was argued that the consumer model of education has dangerous implications for instructional quality if it leads to a reduction in rigor and standards, and it obscures the student's role as a producer of knowledge. Perhaps the faculty at the University of Calgary should continue to resist widescale publication of ratings to avoid pitfalls inherent in trying to please the "consumer student". Faculty concerns and beliefs about SEFs were discussed and some were shown to be misconceptions. However, it appears that student satisfaction can be influenced by such variables as teacher expressiveness, gender, rank of professor, and possibly expected grade in the course. The validity and reliability of SEFs must continue to be discussed in the literature, and faculty may find it in their best interests to be aware of variables other than content coverage that affect student ratings. A new area that may cause initial unreliabilities in student ratings is the integration of technology for teaching and learning. Future research into the outcomes of technology-supported curricula should include careful analysis of the impact of the shifting paradigm from teacher-directed to student-centered instruction on student evaluations of faculty.

References

Abbott, R. D., Wulff, D. H., Nyquist, J. D., Ropp, V. A., & Hess, C. W. (1990). Satisfaction with processes of collecting student opinions about instruction: The student perspective. The Journal of Educational Psychology, 82(2), 201-206.

Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? The Journal of Educational Psychology, 72(1), 107-118.

Abrami, P. C., d'Apollonia, S., Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. The Journal of Educational Psychology, 82(2), 219-231.

Andre, R., & Frost, P. J. (1997). Researchers hooked on teaching: Noted scholars discuss the synergies of teaching and research. Thousand Oaks, CA: Sage Publications, Inc.

Arreola, R. A. (1995). Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system. Bolton, MA: Anker Publishing Company, Inc.

Astin, A. W. (1993). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. American Council on Education, Phoenix, AZ: Oryx Press.

Basow, S. A. & Distenfeld, M. S. (1985). Teacher expressiveness: More important for male teachers than female teachers? The Journal of Educational Psychology, 77(1), 45-52.

Basow, S. A., & Silberg, N. T. (1987).Student evaluations of college professors: Are female and male professors rated differently? The Journal of Educational Psychology, 79(3), 308-314.

Basow, S. A. (1990). Effects of teacher expressiveness: Mediated by teacher sex-typing? The Journal of Educational Psychology, 82(3), 599-602.

Basow, S. A. (1995). Student evaluations of college professors: When gender matters. The Journal of Educational Psychology, 87(4), 656-665.

Beatty, M. J. & Zahn, C. J. (1990). Are student ratings of communication instructors due to "easy" grading practices?: An analysis of teacher credibility and student-reported performance levels. Communication Education, 39, 275-282.

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. The Journal of Educational Psychology, 74(2), 170-179.

Bess, J. L. (1997). Teaching well and liking it: Motivating faculty to teach effectively. Baltimore, MD: Johns Hopkins University Press.

Bryant, J., Comisky, P. W., Crane, J. S., & Zillmann, D. (1980). Relationship between college teachers' use of humor in the classroom and students' evaluations of their teachers. The Journal of Educational Psychology, 72(4), 511-519.

Cahn, S. M. (1978). Scholars who teach: The art of college teaching. Chicago, IL: Nelson-Hall.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 51(3), 281-309.

Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity and usefulness. Review of Educational Research, 41(5), 511-535.

Doyle, Jr., K. O. & Crichton, L. I. (1978). Student, peer, and self evaluations of college instructors. The Journal of Educational Psychology, 70(5), 815-826.

Drucker, A. J., & Remmers, H. H. (1951). Do alumni and students differ in their attitudes toward instructors?. The Journal of Educational Psychology, 42(3), 129-143.

DuCette, J. & Kenney, J. (1982). Do grading standards affect student evaluations of teaching? Some new evidence on an old question. The Journal of Educational Psychology, 74(3), 308-314.

Ehrmann, S. C. (1995). Asking the right questions: What does research tell us about technology and higher learning? Change, 27(2), 20-27.[Online] Available: http://www.learner.org/content/ed/strat/eval/ACPBRightQuestion.html

Ellis, R.(1993). Quality assurance for university teaching. Buckingham: Open University Press.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. Research in Higher Education, 9, 199-242

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. Research in Higher Education, 18(1), 3-124.

Flood, B. J., & Moll, J. K. (1990). The professor business: A teaching primer for faculty. Medford, NJ: Learned Information, Inc.

Frymier, A. B. & Thompson, C. A. (1992). Perceived teacher affinity-seeking in relation to perceived teacher credibility. Communication Education, 41, 388-399.

GFC: Instructor ratings tool approved, decision to release results postponed. (1996, January 27). The University of Calgary Gazette, 26(28).

Gorham, J. & Christophel, D. M. (1990). The relationship of teachers' use of humor in the classroom to immediacy and student learning. Communication Education, 39, 46-62.

Haskell, R. E. (1997). Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century. Education Policy Analysis Archives, 5(6). [Online] Available: http://olam.ed.asu.edu/epaa/v5n6.html (February 12).

Hogan, T. P. (1973). Similarity of student ratings across instructors, courses and time. Research in Higher Education, 1, 149-154.

Holmes, D. S. (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. The Journal of Educational Psychology, 63(2), 130-133.

Howard, G. S. & Maxwell, S. E. (1980). Correlations between student satisfaction and grades: A case of mistaken causation? The Journal l of Educational Psychology, 72(6), 810-820.

Howard, G. S. & Maxwell, S. E. (1982). Do grades contaminate student evaluations of instruction? Research in Higher Education, 16(2), 175-188.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. The Journal of Educational Psychology, 77(2), 187-196.

Jepson, V. (1996a, September 5). The future of teaching evaluations. The University of Calgary Gauntlet, 37(11), A5.

Jepson, V. (1996b, September 19). Who's grading who?: SU proposal to publish teaching evaluations on hold. The University of Calgary Gauntlet, 37(13), A1.

Jepson, V. (1997, March 27). GFC gives go-ahead: Public evaluations approved in principle. The University of Calgary Gauntlet, 37(34), 5.

Johnson, G. R. (1995). First steps to excellence in college teaching. (3rd ed.). Madison, WI: Magna Publications, Inc.

Jones, D. C. (1995). The spirit of teaching excellence. Calgary, AB: Detselig Enterprises, Inc.

Kierstead, D., D'Agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. The Journal of Educational Psychology, 80(3), 342-344.

Kimbal, B. A. (1988). Teaching Undergraduates: Essays from the Lilly Endowment Workshop on Liberal Arts. Buffalo, NY: Prometheus Books.

Kremer, J. F. (1990). Construct validity of multiple measures in teaching, research, and service and reliability of peer ratings. The Journal of Educational Psychology, 82(2), 213-218.

Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. Review of Educational Research, 50(4), 525-544.

Kulik, J. A., & Kulik, C. C. (1987). Review of recent research literature on computer-based instruction. Contemporary Educational Psychology, 12, 222-230.

Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multi-trait, multi-method analysis. The Journal of Educational Psychology, 74(2), 264-279.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. The Journal of Educational Psychology, 76(5), 707-754.

Marsh, H. W. (1994). Weighting for the right criteria in the instructional development and effectiveness assessment (IDEA) system: Global and specific ratings of teaching effectiveness and their relation to course objectives. The Journal of Educational Psychology, 86(4), 631-648.

Marsh, H. W. (1995). Still weighting for the right criteria to validate student evaluations of teaching in the IDEA system. The Journal of Educational Psychology, 87(4), 666-679.

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30(1), 217-251.

Maslow, A. H., & Zimmerman, W. (1956). College teaching ability, scholarly activity and personality. The Journal of Educational Psychology, 47(3), 185-189.

Meier, R. S. & Feldhusen, J. F. (1979). Another look at Dr. Fox: Effect of stated purpose for evaluation, lecturer expressiveness, and density of lecture content on student ratings. The Journal of Educational Psychology, 71(3), 339-345.

McKeachie, W. J. (1978). Teaching tips: A guidebook for the beginning college teacher. (7th ed.). Lexington, MA: D.C. Heath and Company.

McKeachie, W. J., Lin, Y. G., Daugherty, M., Moffett, M. M., Neigler, C., Nork, J., Walz, M., & Baldwin, R. (1980). Using student ratings and consultation to improve instruction. British Journal of Educational Psychology, 50, 168-174.

McKeachie, W. J. (1990). Research on college teaching: The historical background. The Journal of Educational Psychology, 82(2), 189-200.

McMurtry, J. (1991). Education and the market model. Journal of Philosophy of Education, 25(2), 209-217.

Mingle, J. R. (1995). Vision and reality for technology-based delivery systems in post-secondary education. St. Louis, MO: Paper presented at the Governor's Conference on Higher Education. (Eric Document Reproduction Service No. 394 400)

Murray, B. (1997). How important is teaching style to students? American Psychological Association Monitor, 28(5), [Online] Available: http://www.apa.org/monitor/may97/ceci.html (July 20, 1997).

Murray, H. G., Rushton, P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. The Journal of Educational Psychology, 82(2), 250-261.

Myers, D. G. (1997). Professing psychology with passion. In R. J. Sternberg (Ed.), Teaching introductory psychology: survival tips from the experts. (pp. 107-118). Washington, DC: American Psychological Association.

Overall, J. U. & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. The Journal of Educational Psychology, 72(3), 321-325.

Perry, R. P., Abrami, P. C., & Leventhal, L. (1979). Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. The Journal of Educational Psychology, 71(1), 107-116.

Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. Research in Higher Education, 7, 193-205.

Publication of instructor ratings endorsed in principle. (1997, March 24). The University of Calgary Gazette, 26(35).

Questionnaire for rating profs approved. (1997, February 24). The University of Calgary Gazette, 26(32).

Rate the prof?. (1996, October 21). The University of Calgary Gazette, 26(17).

Roblyer, M. D., Edwards, J., and Havriluk, M. A. (1997). Integrating educational technology into teaching. New Jersey: Prentice-Hall.

Ryan, J. J., Anderson, J. A., & Birchler, A. B. (1980). Student evaluation: The faculty responds. Research in Higher Education, 12(4), 317333.

Sandler, B. R. (1991). Women faculty at work in the classroom, or, why it still hurts to be a woman in labor. Communication Education, 40, 6-15.

Schrank, B. (1993). Academic freedom and the inclusive university. Paper presented at the Congress of Learned Societies, Carleton University, Ottawa, Ontario, Jun 5. [Online] Available: http://www.caut.ca (or Contact: Louise Desjardins, FAX 613-820-7244, Canadian Association of University Teachers (CAUT)).

Skierka, L. (1996a, October 3). Oh yes, democracy: Students measure the proposal to publish teaching evaluations. The University of Calgary Gauntlet, 37(15), A1.

Skierka, L. (1997, January 30). Public evaluations postponed: But universal rating system approved in principle. The University of Calgary Gauntlet, 37(27), A1.

Skinner, B. F. (1954). The science of learning and the art of teaching. Harvard Educational Review, 24, 86-97.

Sternberg, R. J. (1997). Teaching introductory psychology: Survival tips from the experts. Washington, DC: American Psychological Association.

Stone, J. E. (1995). Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level. Education Policy Analysis Archives, 3(11). [Online] Available: http://olam.ed.asu.edu/epaa/v3n11.html (July 4, 1997).

Student ratings news. (1997, June). The University of Calgary Faculty Association InterViews, 28(4), 4-5. (Note: Ceci's study cited from The Chronicle of Higher Education).

Violato, C., McDougall, D., & Marini, A. (1992). Educational measurement and evaluation. Dubuque, IA: Kendall/Hunt.

Ware, J. E., Jr., & Williams, R. G. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. Journal of Medical Education, 50, 149-156.

Ware, J. E., Jr., & Williams, R. G. (1977). Discriminant analysis of student ratings as a means for identifying lecturers who differ in enthusiasm or information giving. Educational and Psychological Measurement, 37, 627-639.

Weingartner, R. H. (1993). Undergraduate education: Goals and means. American Council on Education, Phoenix, AZ: Oryx Press.

Widmeyer, W. N. & Loy, J. W. (1988). When you're hot, you're hot! Warm-cold effects in first impressions of persons and teaching effectiveness. The Journal of Educational Psychology, 80(1), 118-121.

Williams, R. G., & Ware, J. E., Jr. (1976). Validity of student ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect. The Journal of Educational Psychology, 68, 48-56.

Williams, R. G., & Ware, J. E., Jr. (1977). An extended visit with Dr. Fox: Validity of student satisfaction with instruction ratings after repeated exposures to a lecturer. American Educational Research Journal, 14, 449-457.